

Avaliação de Desempenho

Tópicos Especiais em Arquitetura de Computadores

Segundo Trabalho Prático

Análise Descritiva de Dados

Observações gerais sobre este trabalho

1. Este é um exercício de programação de análises de dados sobre bases de dados;
2. Você poderá programar em qualquer linguagem ou ambiente que você queira (Python, Shell script, R). Entretanto, a linguagem Python já tem muitas bibliotecas prontas (sci-kit learn, numpy e pandas, por exemplo);
3. Se for aluno de graduação, poderá fazer este trabalho individual, ou em duplas; se for aluno de pós-graduação terá que fazer o trabalho individual;
4. Há alguns materiais disponíveis sobre análise descritiva de dados, a maioria da área de estatística, que podem ser consultados sem qualquer problema;
5. O livro texto dessa disciplina tem muitas dicas importantes e exemplos;
6. Você pode se basear no Jupyter Notebook disponibilizado nos materiais de estudos;
7. A primeira e segunda questão valem **3.0 pontos**. A terceira questão vale **2.0 pontos**. A quarta e quinta questão valem **1.0 ponto**.
8. Considere a base de dados **Fires**, que apresenta dados meteorológicos para previsão de incêndios florestais em um parque, e que é composta por 517 objetos e 13 atributos. Serão utilizados **cinco atributos**, sendo dois categóricos (Mês e Dia) e três numéricos (DMC, DC e ISI).

Exercício 1. Distribuições de frequência

Especificação: Execute os passos para a geração de uma **tabela resumo da distribuição de frequência** de cada um dos cinco atributos (Dia, Mês, DMC, DC e ISI). As colunas de cada tabela serão somente: classe, limite inferior, ponto médio, limite superior, frequência absoluta, frequência relativa, frequência acumulada (valor) e frequência acumulada (percentual).

Entregável: implemente um programa que imprima a tabela resumo da distribuição de frequência para cada um dos cinco atributos. Inclua todos os cálculos necessários no programa. Assuma que a quantidade de **classes** é uma constante ou variável.

Observações:

- Para atributos **categóricos**, não há necessidade de executar **todos** os passos na construção da distribuição de frequência, sendo que a **quantidade de categorias** pode ser utilizada como **número de classes**, bastando determinar a frequência de cada categoria do atributo.
- Qualquer dúvida veja as Tabelas 3.4, 3.6 e 3.7 do livro texto;
- Note que as Tabelas 3.6 e 3.7 apresentam os resumos das distribuições de frequência dos atributos DMC e DC, respectivamente;

- Observe que não é necessário incluir a coluna "Fronteiras".

Exercício 2. Visualização de dados

Especificação: Para cada um dos cinco atributos (Dia, Mês, DMC, DC e ISI), gere os seguintes gráficos: (i) histogramas das frequências relativas e absolutas; (ii) gráfico de polígonos; (iii) gráfico de setores; (iv) gráfico de Pareto; (v) gráfico de dispersão entre cada dois atributos, ou seja, DMC x Dia, DMC x Mês, DMC x DC, DMC x ISI, DC x Dia, DC x Mês, DC x ISI, ISI x Dia, ISI x Mês, Dia x Mês. Além dos gráficos, inclua comentários sobre as informações obtidas pelas análises dos gráficos, deixando claro o gráfico e a informação obtida.

Entregável: implemente um programa que, baseado nos dados, plote os gráficos definidos acima. Os comentários sobre as informações obtidas pelas análises dos gráficos devem ser enviados em um arquivo PDF à parte.

Observações: o livro texto traz exemplos de alguns gráficos sobre alguns dos atributos; o Jupyter Notebook disponibilizado no material de estudos tem exemplos de como construir os gráficos acima.

Exercício 3. Medidas Resumo

Especificação: gere as medidas de tendência central (média, moda (quando for o caso), ponto médio e mediana), dispersão (amplitude, desvio padrão, variância, coeficiente de variação) e forma (assimetria e curtose), para cada um dos cinco atributos (Dia, Mês, DMC, DC e ISI).

Entregável: implemente um programa que calcule e imprima a tabela contendo todas as informações acima.

Observação: a impressão da tabela pode ser em uma única tabela ou em três tabelas separadas.

Exercício 4. Diagrama de Caixa (box plot)

Especificação: imprima os diagramas de caixa (box plot) de cada um dos cinco atributos (Dia, Mês, DMC, DC e ISI).

Entregável: implemente um programa que calcule e imprima os diagramas de caixa (box plot) de cada um dos cinco atributos (Dia, Mês, DMC, DC e ISI). Comente sobre a variabilidade de cada um dos atributos.

Observação: Os comentários sobre a variabilidade de cada atributo devem ser enviados em um arquivo PDF à parte.

Exercício 5. Medidas de Associação

Especificação: imprima os gráficos de dispersão entre os atributos Dia e Mês, DMC e DC, DMC e ISI, e DC e ISI.

Entregável: implemente um programa que calcule e imprima os quatro gráficos de dispersão definidos acima. Baseado nos gráficos de dispersão, comente sobre a associatividade entre cada um dos quatro pares de atributos (correlação positiva, correlação negativa, sem correlação, etc).

Observação: Os comentários sobre a associatividade entre os quatro pares de atributos devem ser enviados em um arquivo PDF à parte.